Transcription Activator-Like Effectors (TALEs) Hybrid Nucleases for Genome Engineering Application

Thesis by

Anjar Tri Wibowo

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

June 2011

EXAMINATION COMMITTEE APPROVALS FORM

The thesis of Anjar Tri Wibowo is approved by the examination committee.

Committee Chairperson Professor Jian Kang Zhu

Committee Member Dr. Liming Xiong

Committee Member Professor Christoph Gehring

ABSTRACT

Transcription Activator-Like Effectors (TALEs) Hybrid Nucleases for Genome Engineering Application

Gene targeting is a powerful genome engineering tool that can be used for a variety of biotechnological applications. Genomic double-strand DNA breaks generated by engineered site-specific nucleases can stimulate gene targeting. Hybrid nucleases are composed of DNA binding module and DNA cleavage module. Zinc Finger Nucleases were used to generate double-strand DNA breaks but it suffers from failures and lack of reproducibility. The transcription activator–like effectors (TALEs) from plant pathogenic Xanthomonas contain a unique type of DNA-binding domain that bind specific DNA targets. The purpose of this study is to generate novel sequence specific nucleases by fusing a de novo engineered Hax3 TALE-based DNA binding domain to a *FokI* cleavage domain. Our data show that the de novo engineered TALE nuclease can bind to its target sequence and create double-strand DNA breaks *in vitro*. We also show that the de novo engineered TALE nuclease is capable of generating double-strand DNA breaks in its target sequence *in vivo*, when transiently expressed in *Nicotiana benthamiana* leaves. In conclusion, our data demonstrate that TALE-based hybrid nucleases can be tailored to bind a user-selected DNA sequence and generate site-specific genomic double-strand DNA breaks. TALE-based hybrid nucleases hold much promise as powerful molecular tools for gene targeting applications.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my committee chair Professor Jian-Kang Zhu for his generous guidance and support throughout this research. I would also like to thank the committee members Dr. Liming Xiong and Professor Christoph Gehring for their time and effort in reviewing this work and for their valuable feedback.

My sincere thanks go to Dr. Magdy Mahfouz for his constant support, assistance and valuable suggestions. My appreciation also goes to my friends and colleagues at the “Genome Engineering Research Group”, Dr. Lixin Li, Mr. Md. Shamimuzzaman, Ms. Xiaoyun Fang, Dr. Mohammad Suhail Khan, Ms. Yveline Pailles, Ms. Agnieszka Piatek, Mr. Marek Piatek and Dr. Shujun Yang, for their help and valuable advice. I also want to extend my gratitude to all the Professors, students, researchers and staff at the Plant Stress Genomics Research Center, King Abdullah University of Science and Technology, for their help and support in all respects during the completion of the project.

Finally, my gratitude is extended to my parents and fiancée for their encouragement, love and patience. For that, I wish to say a heartfelt thank you.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AD	activation domain
DSB	double-strand DNA breaks
EBE	effector binding element
EMSA	Electrophoretic Mobility Shift Assay
NLS	nuclear localization signal
NHEJ	Non-Homologous End Joining
RVD	repeat variable diresidue
TALE	transcription activator-like effectors
UPA	upregulated by AvrBs3
VAGE	virus-aided gene expression
ZFN	zinc finger nuclease

LIST OF SYMBOLS

Symbols for amino acids:

D	aspartic acid
G	glycine
H	histidine
I	isoleucine
K	lysine
N	asparagine
S	serine

Symbols for nucleotides:

A	adenine
G	guanine
T	thymine
C	cytosine

LIST OF ILLUSTRATIONS

I. INTRODUCTION

For centuries, humans have made efforts to secure the food supply by selecting crop varieties of higher yield and resistance to pathogens and stress cues. These selection methods have developed over time to include genetics based breeding programs and genetic engineering. In 1970s, the advent of restriction enzymes has marked the beginning of the genetic engineering where scientists were able to cut and ligate pieces of DNA. In 1980s, scientists were able for the first time to transform plants with foreign genes, ushering in an exciting field of plant genetic engineering. Genetic engineering technology allows us to directly manipulate the genome of an organism without going through breeding and random mutagenesis. Genetic engineering, however, suffers from two major drawbacks, the insertion of the gene of interest into the genome is random and the inserted gene could affect other endogenous genes resulting in unwanted effects (Gepts, 2002; Jauhar, 2001; Wang et al., 2009; Zhong, 2001).

In the last decade, researchers have begun to develop the next generation technology called “genome engineering”. This technology allows a highly efficient, targeted and precise modification of the genome of an organism. The application of this technology would allow a robust and efficient addition, deletion, activation and inactivation of genes with high precision and specificity. The ability to manipulate the genome in such ways would accelerate the trait development

process and expand the range of traits and plant species amenable to genetic modification (Durai, 2005; Wirth et al., 2007).

Genome engineering tools include chimeric proteins that are usually composed of a DNA binding module and an effectors domain with the ability to activate or repress transcription and cleave or modify the DNA of the target gene (Collins, 2003; Urnov et al., 2010). The type IIs restriction endonuclease *FokI* isolated from *Flavobacterium okeanokoites* is arranged in a modular structure consisting of N-terminal DNA recognition domain and a C-terminal non-specific DNA cleavage domain (Durai et al., 2005). DNA cleavage by *FokI* requires the dimerization of *FokI* and the occurrence of magnesium ions. When *FokI* recognition domain bound to duplex DNA at non-palindromic recognition sequence 5′-GGATG-3:5'-CATCC-3', DNA cleavage domain will be activated and cleave DNA non-specifically at 9 bp and13 bp downstream of the recognition site (Bitinaite et al., 1998; Wah et al., 1998). The modular nature of *FokI* has made it an ideal candidate for the construction of hybrid endonucleases with novel sequence specificities, by combining cleavage domain of *FokI* to variety of DNA-binding domains that could recognize longer DNA sequences such as the zinc fingers or TALEs (Gersbach et al., 2010; Gordley et al., 2007).

Zinc-finger nucleases (ZFNs) are artificial hybrid proteins generated by fusing a zinc finger DNA-binding domain to a DNA-cleavage domain of *FokI* endonuclease (Pabo et al., 2001; Vanamee et al., 2007). Zinc finger modules can

be engineered to recognize a particular DNA sequence both *in vitro* and *in vivo* which enables ZFN to induce Double-stranded breaks (DSBs) at specific genomic locus (Cathomen and Joung, 2008; Kim et al., 1996). DSB at a specific genomic locus can be subsequently repaired by the NHEJ (Non-Homologous End Joining) pathway. Imprecise repair by NHEJ leads to occasional loss or gain of the genetic information in form of small insertions or deletions, which could disrupt the gene function (Ramalingam et al., 2010; Wang et al., 2003). Zinc Finger Nucleases (ZFNs) have been used to manipulate the genome of various organisms, including zebra fish (Doyon et al., 2008; Thermes et al., 2002), yeast (Epinat et al., 2003; Li et al., 2011), plants (Chiurazzi et al., 1996; Lloyd, 2005), and mammals (Epinat et al., 2003; Miller et al., 2010). However, the process of designing and creating sequence-specific DNA-binding proteins which are based on zinc fingers remains labor intensive, difficult, expensive and less reproducible (Miller et al., 2010; Ramirez et al., 2008). There is a pressing need for a DNA binding module with extreme specificity that can be adaptable for any user selected target sequence. Recently, scientists developed sequence-specific DNA-binding proteins based on transcription activator-like effectors (TALE) of the pathogen plant *Xanthomonas sp*.

Phytopathogenic bacteria of the genus *Xanthomonas* inject TALE via the type III secretion system. TALE- related proteins are also found in *Ralstonia solanacearum* (Heuer et al., 2007; Yang and White, 2004). These TALEs act as virulence factors, translocated to the nucleus and function as transcription factors

by binding to their DNA target in the promoter region and reprogram the gene expression of the host (Boch and Bonas, 2010; Gürlebeck et al., 2006). The *avrBs3* gene from the pepper and tomato pathogen *Xanthomonas campestris pv. vesicatoria* was the first TALE to be isolated and identified (Minsavage et al., 1990). The most distinct structural feature of TALE is the presence of the tandem repeat central domain. The number of repeats varies from one TALE to another but they range from 1.5 to 33.5 repeats (Boch and Bonas, 2010). The repeat central domain of the *avrBs3* gene consists of 17.5 of 102-bp long DNA repeats, each repeat encodes 34 amino acids (Bonas et al., 1989). After the discovery of the *avrBs3* gene, some homologous genes including *avrXa5, avrXa7,* and *avrXa10* were isolated and identified from the rice pathogen *Xanthomonas oryzae* pv. *Oryzae*. These three proteins exhibit the same structural features of *avrBs3,* they have central domain tandem repeats similar to that of *avrBs3,* albeit with varied number, length, and arrangement of repeats. Structural comparison between AvrBs3 and AvrXa10 first showed that the polymorphism between TALE central domain repeats occurs in the amino acid number 12 and 13 of each repeat, known as the repeat variable diresidue (RVD) (Hopkins et al., 1992). The target of the AvrBs3 in plants was identified in pepper and referred to as *UPA* (upregulated by AvrBs3) (Marois et al., 2002). *UPA* box, a highly conserved promoter region was found in some UPA genes, and several studies showed that AvrBs3 may directly bind to the UPA box. AvrBs3 can bind specifically to target sequences, both *in-vitro* and *in-vivo* (Kay et al., 2007; Kay et al., 2009).

Since the discovery of AvrBs3 family of TALEs, several studies have focused on the contribution of the tandem repeats of the central domain to dictate the specificity of TALEs to the DNA binding box in the promoter region of the target gene. Two major studies employing experimental and bioinformatics methods deciphered the code of TALE binding to the DNA target. These two studies observed that the length of the UPA box is more or less similar to the number of AvrBs3 central tandem repeats. This observation led to testing the hypothesis that each repeat binds to one nucleotide in the DNA target sequence (Boch et al., 2009; Moscou and Bogdanove, 2009).

Moscou and Bogdanove showed that the association between various RVDs of TALEs and DNA nucleotides of the target sequence does not occur at random. Based on the association of 382 RVD–nucleotide studied, they observed that histidine and aspartic acid (HD) is the most common RVD to be found and is strongly associated with nucleotide C. Other most common RVDs are asparagines and glycine (NG), which is associated with nucleotide T, and asparagine and isoleucine (NI) which is associated with nucleotide A. Two asparagines (NN) is the next most frequent RVD to be found and is associated with nucleotide G, but is also found to associated with nucleotide A. Other RVDs also frequently found are asparagine and serine (NS) which was found to be associated with G, A, or C, and the last one is histidine and glycine (HG) of which association resembles that of NG. Several other RVDs were also found, but with very rare occurrence (Fig. 1)(Moscou and Bogdanove, 2009).

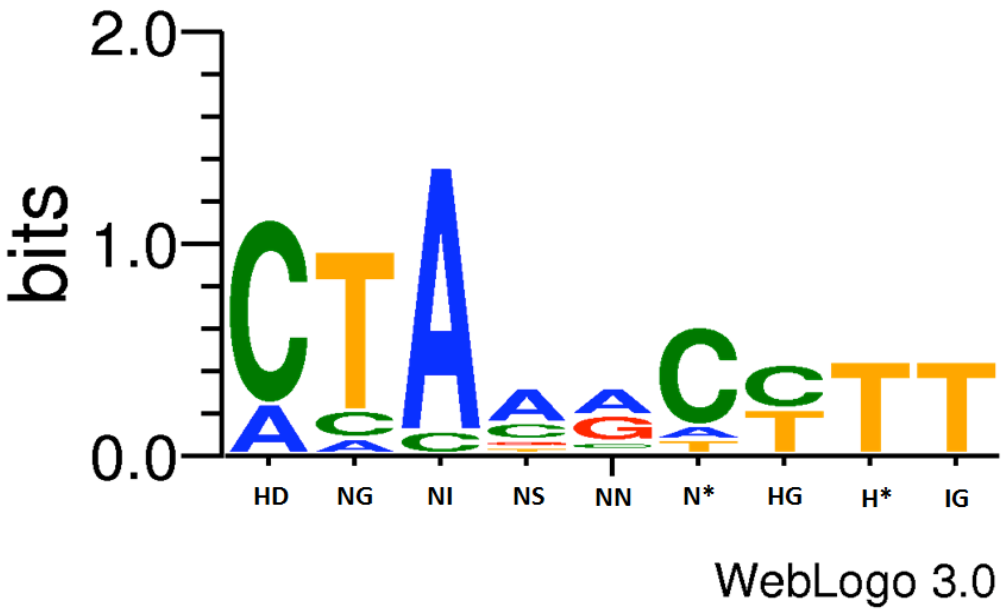


Fig. 1. TALE - DNA recognition code. Letter size represents association frequency between each nucleotide to particular RVD (Moscou and Bogdanove, 2009).

Association between various RVDs of TALE with DNA nucleotides of the target sequence was also proved experimentally. Boch *et al*. deciphered the TALE code by comparing RVDs of AvrBs3 to UPA box sequence, in which they observed that UPA box has the base pair length similar to the number of repeats in AvrBs3 central domain. Based on this observation, they predicted the target sequences for Hax2, Hax3, and Hax4 from *X. campestris pv. Armoraciae* and generate artificial promoters that match these three TALEs. Using transient expression assay they demonstrated that artificial promoters containing target sequence for Hax2, Hax3, or Hax4 were only induced by the specific corresponding effector. Further, to showed that each RVD are recognize specific base pair, they permutated the Hax4 target sequence (Hax4 box). GUS assay demonstrated that permutation of nucleotides A, T, or C on Hax4 box which corresponds to RVDs

NI, NG, and HD respectively, abolished the GUS activity, whereas for NS, permutations had not effect. In addition, they constructed ArtX1, an artificial TALE which include RVD NN. Permutations on ArtX1 box demonstrated that NN corresponds with A or G. These experiments verified the correspondence between each RVD and a DNA nucleotide of the target, where HD corresponds to nucleotide C; NG corresponds to nucleotide T; NI corresponds to nucleotide A; NS corresponds to nucleotides A, C, or G; and NN corresponds to nucleotide A or G. They also concluded that the conserved T preceding the DNA target box was crucial for binding and might interact with repeat 0, an unidentified repeat (Boch et al., 2009). In a recent study, Morbitzer *et al*. added that RVD asparagines and lysine (NK) mediates specific interaction with nucleotide G (Morbitzer et al., 2010). Overall, the experimental evidence from Boch *et al.* and the bioinformatic evidence from Moscou and Bogdanove have resulted in deciphering the TALE-DNA binding code.

The main constraint of the use of the DNA binding domain of TALE is the difficulty of constructing novel TALEs routinely because the DNA binding domain of TALE consists of a large number of nearly identical tandem-arranged repeats. This highly homologous repetitive structure might lead to plasmid instability during cloning rendering the process of constructing customized TALEs relatively difficult. Gene synthesis of TALEs with novel repeat specificity is possible but relatively expensive and time consuming (Cermak et al., 2011; Li et al., 2011; Morbitzer et al., 2011; Zhang, 2011).

Recently several groups have reported the construction of novel TALEs with customized specificity. Usually the construction of new TALEs stars with codon optimization to facilitate their expression in specific organism. Miler *et al*. reported the construction of customized TALE that can recognize the human endogenous genes CCR5 and NTF3 using infusion cloning of several overlapping repeats fragments (Miller et al., 2010). While Zhang *et al.* developed a hierarchical ligation-based method to overcome the difficulty of TALEs construction with tandem repeat domains. They used a method similar to the Golden Gate cloning for multiple DNA fragment ligation. In order to assemble each individual repeats monomer in a specific order, they modified the DNA sequence at the junction between each pair of repeat monomer. Using a variety of different codons, at each junction they designed unique 4-bp sticky-end ligation adapters. Using this method, four repeat monomers can be assembled directly into 4-mer intermediary arrays. Three 4-mer intermediary arrays can be combined directly to produce a 12-mer intermediary arrays which then ligated to a backbone vector to produce a full-length TALE with 12 RVDs (Fig 2.)(Zhang, 2011). Cermak *et al.* applied the similar system, they reported that the construction process of a novel TALE can be done within 5 days, consisting of 2 stages: (1) constructing several repeat monomers into intermediary arrays which consist of 1-10 repeats and (2) combining intermediary arrays with a backbone vector to produce a full-length TALE (Cermak et al., 2011).

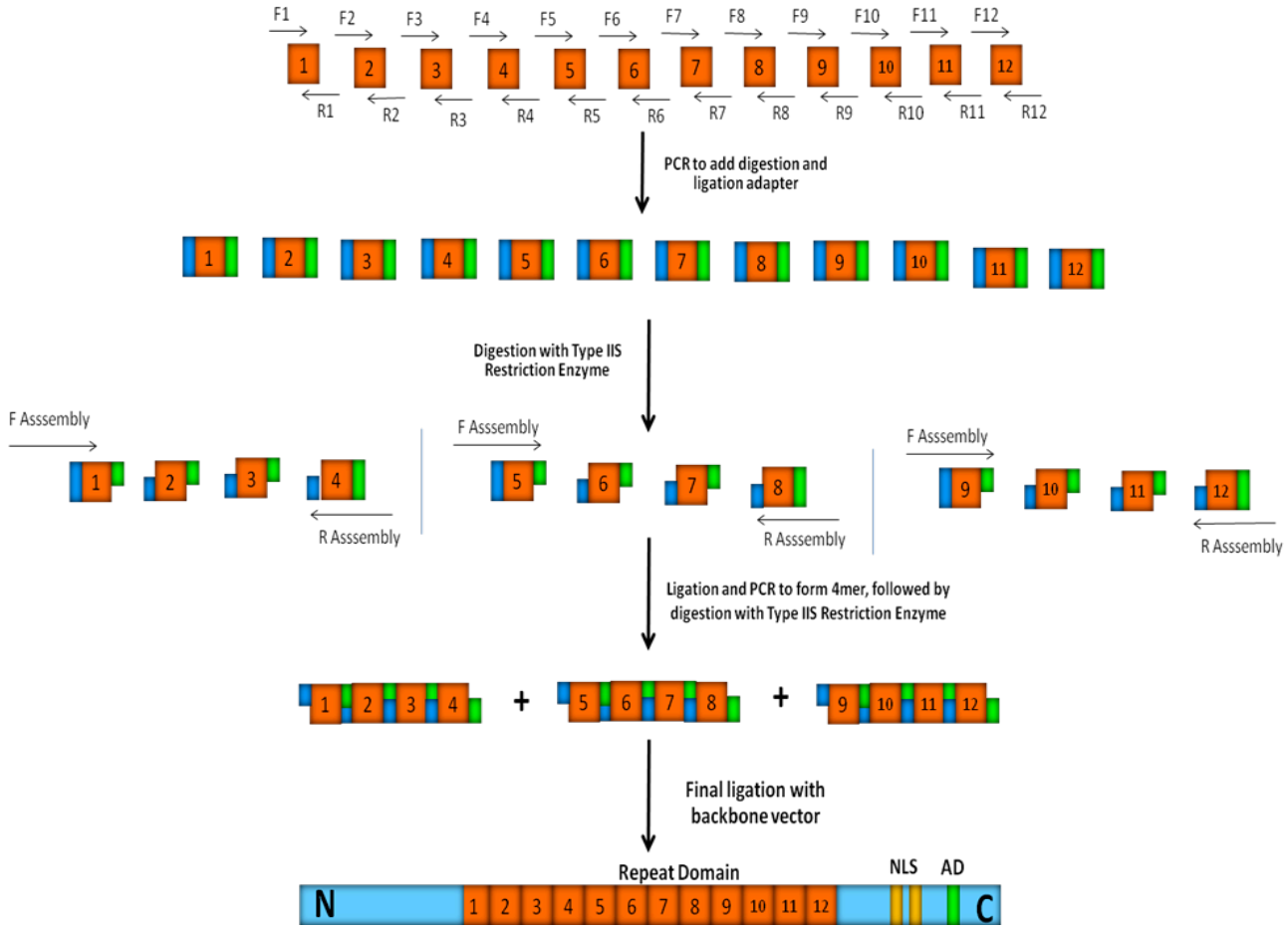


Fig. 2. Assembly of a customized TALEs in serial ligations with type IIs restriction endonuclease. Type IIS restriction endonucleases can cut DNA outside of its recognition site, producing unique 4-bp sticky-end overhangs. The cloning process can be accelerated because the digesting and ligating processes are performed in a single reaction because the appropriate assembly of repeat monomers will eliminate the enzyme recognition site (Zhang, 2011).

The protocols developed by Cermak *et al.* and Zhang *et al.* involves several times of amplifications and ligations to construct each individual repeat. Morbitzer *et al.* developed less laborious protocols by combining a modular assembly based method with Golden Gate cloning method that allows the construction of

TALEs from multiple repeats fragments by using type IIs restriction enzymes in two sequential cut-ligation reactions. Using the method, each individual repeat of monomer can be constructed specifically to assemble functional repeat arrays that can recognize user-predefined target sequence (Morbitzer et al., 2011).

Similarly, Li *et al.* and Weber *et al.* also reported the development of modular assembly-based construction method. Li *et al.* used four repeats monomers with RVD NI, NG, NN and HD to create 48 ready-to-use modules that can be constructed into functional repeat arrays with varying number of repeat units up to a maximum of 23 repeats. The construction of each repeat module can be performed in a controlled order, based on a specific DNA target sequence that the user desired (Li et al., 2011). Weber *et al.* developed a strategy to construct novel TALEs using 17.5 repeat-containing AvrBs3 TALE as a scaffold. They create 68 ready-to-use repeat modules, where DNA binding domain for any 17 base pair target sequence of user choice can be constructed by assembling selected modules in two cloning steps using the Golden Gate cloning method that allows directional and continuous assembly of multiple DNA fragment. The modular assembly-based methods above are quite simple, fast, and inexpensive and in the future will allow high-throughput TALEs synthesis (Weber et al., 2011).

DNA binding domain of TALE and its simple correspondence with a base DNA target opens an array of potential biotechnologies applications. Several studies

have reported the success of designing custom TALEs that can recognize specific DNA sequences and regulate the transcription of diverse endogenous genes of the target genome in various cell types including yeast, plants and mammals (Miller et al., 2010; Morbitzer et al., 2010; Zhang, 2011). TALEs can be combined with various effectors domain, akin to ZFNs, we can create a transcription activator-like effector nucleases by combining the DNA binding domain of TALE with a non-specific cleavage domain from the type II restriction endonuclease *FokI*. Recently, Cermak *et al.* demonstrated the success of the use of novel customized TALE for inducing DSB and mutagenesis of the ADH1 gene in Arabidopsis protoplast. They show that a custom TALE can create DSB on the recognition site and lead to mutation in the form of 4-15 bp deletions at the spacer locus. This mutation is attributable to the imprecise repair of the cleaved chromosomes by NHEJ process (Cermak et al., 2011).

In mammalian system, Miller *et al.* demonstrated the use of TALE in activating and modifying endogenous genes in cultured human cells. The natural 13-repeats TALE protein was replaced with 18 repeats designed to recognize 18-bp section of the promoter region of the endogenous human gene NTF3 (neurotrophin-3). The 18 repeat TALE proteins were combined with a transcriptional activator, VP16. They demonstrated that the hybrid protein can activate the transcription of NTF3 and increase NTF3 expression level up to twentyfold. Then, the VP16 domain was replaced with *FokI* nuclease. Similar to various other studies on yeast and Arabidopsis, the hybrid nuclease protein is

able to induced DSB at specific target sequence on the human genes *NTF3* and *CCR5*, which followed by NHEJ-mediated genome modification in the form of small deletions. They observed that up to 9% of the treated cells carried mutations, when the NTF3 gene was targeted by a custom TALE nuclease. On the CCR5 gene, more than 20% mutated alleles of CCR5 gene were observed. In addition, Miller *et al.* also demonstrated the use of TALE for gene editing through homology directed repair (HDR). Custom TALE nucleases and donor DNA fragments constructed to transfer 46-bp were inserted to K562 cells. They succeeded in showing an efficient and precise editing, in which 16% of alleles having the expected insertion (Miller et al., 2010).

Because of the relatively simple structure of the DNA-binding domain of TALE and the potential of using TALE to identify long DNA sequences, engineered TALE can be generated reproducibly and can overcome the limitations of and ZFN technologies. Engineered TALE may significantly advance the genome modification technologies and be a powerful tool for agricultural biotechnology or gene therapy. Here we reported the generation of novel sequence specific nucleases by fusing engineered Hax3 TALE-based DNA binding domain to a *FokI* cleavage domain. We showed that our de novo engineered TALE nuclease can bind to its target and create DSB *in vitro*. We also showed its ability to create a DSB in its target sequence *in vivo*, when transiently expressed in *N. benthamiana* leaves.

II. MATERIAL AND METHODS

dHax3 TALE Assembly and Vector Construction

Optimization of *Hax3* cDNA is performed to reduce the homology of central repeats and to decrease the value of GC content. *dHax3* and *dHax3.N* cDNAs used in this study was synthesized by the Bio BlueHeron in a pUC19 vector. The cDNAs were amplified and cloned in the pENTR/D vector to produce pENTR/*dHax3* and pENTR/*dHax3.N* gateway entry clones. The resulting clones were confirmed by sequencing (Appendix 1). LR recombination was performed between the entry clones and the gateway-compatible pET32a expression vector, to produce pET32a.*dHax3* and pET32a.*dHax3*, as described in the instruction manual from the manufacturer. The expression clones were transformed into *Escherichia coli* BL21. The protein expression was induced by adding β-D-1thiogalactopyranoside (IPTG) in the culture medium and incubating the *E. coli* culture at 25° C for 5 hours. The His.dHax3, His.dHax3.N, TRX. dHax3, and TRX.dHax3.N proteins were purified using Qiagen Ni-NTA agarose resin, as described in the instruction manual from the manufacturer. To produce the overexpression clones pK2GW7/dHax3 and pK2GW7/dHax3.N, LR reactions were performed between the pENTR/dHax3 and pENTR/dHax3.N entry clones and the pK2GW7 gateway compatible binary vector. The expression clones produced were subsequently transformed into *A. tumefaciens* GV3101 for use in transient expression in tobacco leaves.

Electrophoretic mobility shift assay (EMSA)

TRX.dHax3 and TRX.dHax3.N proteins were purified from *E. coli* BL21 as previously described. The protein concentration was measured with a protein assay kit (BioRad). For binding studies, 5 'biotinlabeled or nonlabeled oligonucleotides were annealed and used as probes. Using the Lightshift Chemiluminscent EMSA kit (Pierce), all of the EMSA reactions were performed as described in the instruction manual from the manufacturer. 12 mM Tris·HCL, 60 mM KCl, 1 mM DTT, 2.5% glycerol, 5 mM MgCl2, 50 ng/μL poly(dI.dC), 0.05% NonidetP-40, 0.2 mM EDTA, 35 fmol biotin-labeled DNA, 10 pmol unlabeled DNA, and 30–480 fmol TRX.His fusion proteins were contained in EMSA binding-reaction buffer. For 10 minutes, the EMSA binding reactions were reserved on ice before the biotin-labeled probe was added. Using 8% Tris/Borate/EDTA (TBE) native ready-made gels from Invitrogen, gel electrophoresis was conducted. Blotting was performed on a nylon membrane with positive charges, and then it was cross-linked using CL-1000 UV cross-linker for 30 s.

dHax.3N *in vitro* DNA Cleavage Analysis

dHax.3N EBEs were cloned in a tail to tail orientation in sense and antisense strands, respectively. The EBEs cloning was achieved by the synthesis of the EBEs complementary ssDNA with 3′A overhangs to assist TA cloning. The EBE ssDNA was recombined and replicated into the TOPO TA cloning vector (Invitrogen), and the EBE clones were identified by sequencing. The

pCRII/EBE6, pCRII/EBE10, and pCRII/EBE16 clones were digested to completion by NcoI, and 500 ng linearized EBE clones were utilized in the digestion reactions with Trx.His.dHax.3N-purified protein. The digestion reactions were conducted in a buffer containing 20 mM Tris·HCl (pH 7.4), 5 mM MgCl2, 50 mM KCl, 5% glycerol, and 0.5 mg/mL BSA. They were then incubated for 30 minutes at 37°C and further analyzed by DNA gel electrophoresis. To study the effects of mutation in the DNA target on the cleavage activity dHax3.N enzyme, we constructed two mutant clones. First, we removed the T nucleotide that matches with repeat 0, and second by substituting the last T that matches with repeat 11.5 with A.

A. tumefaciens-Mediated Transient Expression Assay in *N. benthemiana* leaves

We used transient assays in *N. benthamiana* leaves, to study the activity of dHax3.N *in vivo*. The dHax3.N target was designed to include the combination of EBE-stop codon.spacer-EBE fused in frame to uidA cDNA and replicated in a binary vector to create pK2GW7 (35S:: EBE-TGA. Spacer.-EBE-GUS). The EBE-TGA.spacer-GUS clone was created by amplifying the uidA cDNA using the EBE-TGA-GUS-F and GUS-R primers. The PCR product was cloned into the pENTR/D vector and subcloned by LR reaction in the pK2GW7 overexpression vector. The pK2GW7/EBE-TGA.spacer-EBE-uidA and pK2GW7/dHax3.N were transformed into *A. tumefaciens* GV3101 and co-infiltrated or infiltrated

separately into tobacco leaves, and the infiltrated discs were collected 48 hpi. The GUS qualitative assay was conducted by immersing leaf discs in GUS-staining solution (which contains 10 mM sodium phosphate, pH 7, 10 mM EDTA, 0.1% Triton X-100, 0.1% 5- bromo-4-chloro-3-indolyl-β-D-glucuronide, 1 mM potassium ferricyanide, 1 mM potassium ferrocyanide) at 37 °C for 24 h. The discs were cleared in ethanol.

III. RESULTS

De Novo Hax3 TALE-Based Hybrid Nuclease Design and Assembly

We selected Hax3 effector of TALEs protein family as the basis for the construction of our customized TALE-based nuclease, because Hax3 protein is relatively smaller compared to other members of TALEs family. Hax3 has 11.5 repeats in the central DNA binding region; repeats of that size are adequate for specific and strong binding as well as for activating target promoter (Boch et al., 2009)(Fig. 3A). TALE genes are usually constructed by using PCR-based method, and to modify the RVDs of TALE PCR mutagenesis-based method is generally used. The sequence of natural Hax3 cDNA (GenBank accession no. AY993938.1) has a relatively high overall GC content i.e. 65% and its 11.5 tandemly arranged repeats in the central DNA binding region are almost identical one to another which makes PCR mutagenesis of these repeats challenging. In order to overcome these limitations, we created codon-optimized versions of the original *Hax3* cDNA, herein defined to as dHax3. dHax3 are codon-optimized for *in-planta* expression, which has a GC content of only 47% and show less homology between the 102-bp repeats compared to the original *Hax3* cDNA.

Natural Hax3 TALE can attach to the target DNA with sequence TACACCCAAACAT (Boch et al., 2009). We modified the RVDs of the original Hax3, so that the dHax3 we constructed can attach to the 12-bp sequence TCCCTTTATCTCT in the *RD29A* promoter (Yamaguchi-Shinozaki and

Shinozaki, 1993)(Fig. 3B). It should be noted that the target sequence begins with nucleotide T at position 0. It was previously reported that nucleotide T at position 0 which found in all TALE target sites, contributes to TALE mediated activation of the respective promoters of target genes, and might also be crucial in the binding process of TALE to the target sequence.

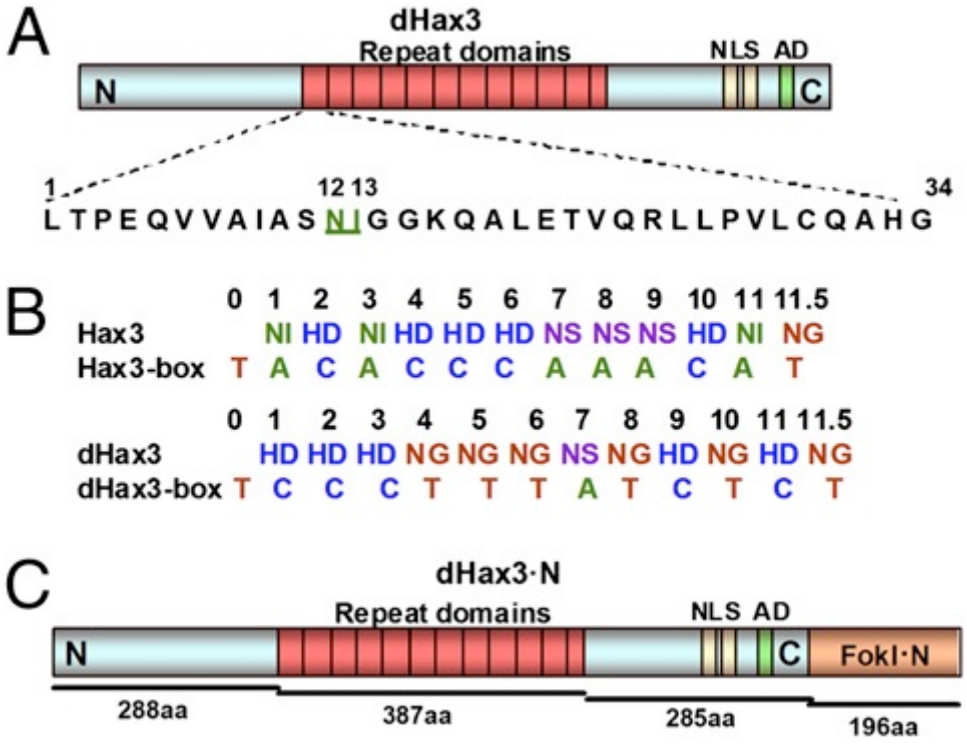


Fig. 3. Schematics of a Hax3 and dHax3.N proteins and their DNA target specificity. **(A)** Hax3 TALE has a central DNA binding domain with 11.5 repeats units (repeat domains, shown in red), in which each repeats unit consists of 34 aa, representation of the first repeat is shown below, with RVD at position 12 and 13 shown in green. A nuclear localization motifs (NLS) shown in yellow, and a transcription activation domain (AD), at the C-terminus, shown in green. **(B)** RVDs composition of the 11.5 repeat units of Hax3 and dHax3 and its corresponding 12-bp DNA-binding target. (C) Schematic representation of the dHax3.N with the *FokI* cleavage domain at the C-terminus of dHax3.

A hybrid dHax3 nuclease is constructed by fusing in frame the DNA sequence encoding the full-length dHax3 sequence with DNA sequence encoding 196 aa of the *FokI* C-terminal cleavage domain. The type II restriction endonuclease *FokI* was selected because several previous studies have reported non-specific catalytic activity of *FokI* when combined with other DNA-binding domains, such as DNA binding domain of zinc finger proteins (Mani et al., 2005). The hybrid dHax3 nuclease, henceforth referred to as dHax3.N, has 1,157 aa residues and about 124 kDa molecular mass (Fig. 3C). In the design of dHax3.N, dHax3 was designed to provide DNA binding specificity, while the cleavage domain of *FokI* is to provide nuclease activity. dHax3.N was synthesized in fragments which then fused together and cloned into a pUC19 vector and then, into a pENTR/D gateway vector; sequencing was used to confirmed the constructs. The dHax3 and dHax3.N clones were then subcloned into the pET32a expression vector to produce thioredoxin.6His fusion proteins (Trx.His.dHax3 and Trx.His.dHax3.N). It was shown that the bacterially expressed Trx.6His.dHax3 and Trx.6His.dHax3.N proteins migrated in SDS protein gel electrophoresis to their estimated sizes (Appendix 2).

DNA-Binding Activities of dHax3 and dHax3.N Hybrid Nucleases *in vitro* and Activation of Target Gene Expression in Vivo.

The Trx.6his-tagged dHax3 and dHax3.N proteins were purified from *E. coli* and then tested for their DNA-binding specificity to the 12-bp EBE *in vitro*. The ability

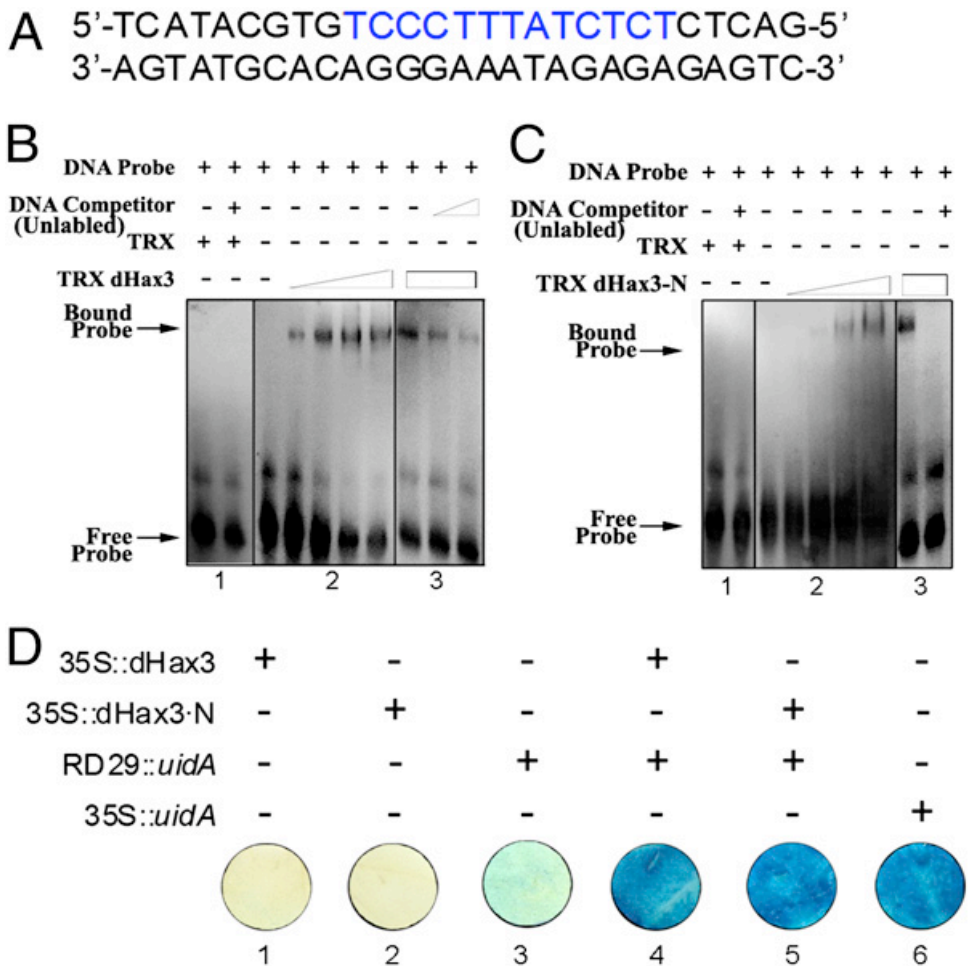


Fig. 4. Binding specificity of dHax3 and dHax3.N to their target DNA and *in vivo* transcriptional activation. **(A)** Double-stranded oligonucleotides probe used in the EMSA, containing dHax3 and dHax3.N EBE target sequence (shown in blue) **(B)** EMSA demonstrating specificities of dHax3 and dHax3.N binding to the EBE target sequence; Panel 1 thioredoxin tag can not bind to biotinlabeled probe. Panel 2 dHax3 protein bind to biotinlabeled probe sequence in a concentration-dependent manner. Panel 3 unlabeled probes competitively reduced the binding between dHax3 protein and biotinlabeled probe **(C)** EMSA demonstrating specificities of dHax3 and dHax3.N binding to the EBE target sequence; Panel 1 thioredoxin tag can not bind to biotinlabeled probe. Panel 2 dHax3.N protein bind to biotinlabeled probe sequence in a concentration-dependent manner. Panel 3 the competition between biotinlabeled probe and unlabeled probe to bind to dHax3.N **(D)** dHax3 and dHax3.N transcriptionally activates RD29A promoter in vivo. Panels 1 and 2 acts as negative control, shows that the addition of dHax3 or dHax3.N only does not trigger uidA activity. Panel 3 shows the background activity of RD29A promoter. Panels 4 and 5 shows the transcriptional activation of RD29A promoter by dHax3 and dHax3.N, respectively. Panel 6 A positive control of 35S::uidA.. In the three times the experiment was performed, similar results were achieved; each used 60 plants (10 plants per panel).

of purified Trx.6his-tagged dHax3 and dHax3.N to attach to the DNA target *in vitro* was tested with EMSA. Both proteins were used in EMSA along with biotin labeled double-stranded oligonucleotides, containing the EBE target sequence (Fig. 4A).

The EMSA demonstrated that Trx.6His.dHax3 and Trx.6His.dHax3.N proteins can attach specifically to double-stranded oligonucleotides containing their EBE target sequence. dHax3 and dHax3.N bindings to labeled double-stranded oligonucleotides competitively reduced with the addition of unlabeled double-stranded oligonucleotides of the same sequence (Fig. 4B and C). Data from this study also shows that thioredoxin tag did not affect the binding of the dHax3 and dHax3.N proteins to the EBE target sequence.

Aside from having DNA binding domain, TALEs also have C-terminal transcription activation domain. We have hypothesized that the combination of *FokI* C-terminal cleavage domain on the C-terminus of dHax3 will not affect the activation domain of dHax3 and the activation domain will remain functional. To investigate whether dHax3 and dHax3.N proteins can attach to target sequence *in vivo* and activate gene transcription, we subcloned dHax3 and dHax3.N in binary vector under control of the strong CaMV 35S promoter. The 35S::dHax3 and 35S::dHax3.N constructs were transformed into *A. tumefaciens* and delivered by injection into *N. benthamiana* leaves for transient expression

assays. A 300-bp RD29A promoter was cloned upstream of the uidA cDNA in the pK2GWFS7 binary vector (Karimi et al., 2007). The 35S::dHax3 and 35S::dHax3.N containing Agrobacteria were co-infiltrated in *N. benthemiana* leaves with Agrobacterium that contained RD29A::uidA. Leaf discs of the infiltrated regions were collected 24 h post-infection (hpi) and dyed with beta-glucuronidase (GUS) staining buffer to observe the activity of the UidA reporter. The data demonstrated that dHax3 and dHax3.N can bid to and activate the RD29A promoter (Fig. 4D).

DNA cleavage activities of dHax3.N Hybrid Nuclease

Next we wanted to test our de novo engineered TALE nuclease protein ability to cut the target DNA sequence *in vitro*. We constructed plasmid containing two EBE target sequences in a tail-to-tail orientation separated by spacers with different lengths. To help the data analysis, we incorporated restriction sites for endonuclease SmaI in the spacer regions. We constructed three clones pCRII/EBE6, pCRII/EBE10, and pCRII/EBE16, each of which has a spacer region with 6-bp, 10-bp, and 16-bp long (Fig. 5A and B). We use thiroedoxin-tagged dHax3.N protein purified from bacteria for digestion analysis. First, to linearize pCRII/EBE clones, we cut the pCRII/EBE clones at a unique restriction site for NcoI or XcmI. Further, the plasmid is gel-purified and used as a substrate to test dHax3.N cleavage ability. In our digestion analysis, we used various buffers, incubated together with purified dHax3.N protein and linearized

pCRII/EBE clones, with 30 minutes incubation time at 37° C. We did not observe any cleavage activity when using the EMSA DNA-binding buffer (12 mM Tris·HCL, 60 mM KCl, 2.5% glycerol, 5 mM MgCl2), whereas when using (20 mM Tris·acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9) we only observed slight cleavage activity. Cleavage activity of dHax3.N is highest observed when we used the reaction buffer containing 20 mM Tris·HCl (pH 7.4), 5 mM MgCl2, 50 mM KCl, 5% glycerol, and 0.5 mg/mL BSA, where dHax3.N cut linearized pCRII/EBE plasmid into two fragments of the expected size (1.6 and 2.4 kb).

In this experiment we used clones with various spacer lengths between dHax3.N binding sites to identify the most optimal spacer lengths to support *FokI* dimerization and nuclease activity. We observed that dHax3.N does not show cleavage activity when incubated with EBE clone with a 6-bp spacer-length clone (pCRII/EBE6) while a little cleavage activity was observed when incubated with EBE clone with a 10-bp spacer-length (pCRII/EBE10). Robust and highest cleavage activity was observed when dHax3.N was incubated with EBE clone with a 16-bp spacer-length (pCRII / EBE16), in which we observed that the pCRII/ EBE16 was cut into an expected cleavage pattern (Fig. 5C). Our results correspond to several other reports which mention that the length of spacer region is important for TALE Nuclease cleavage activity.

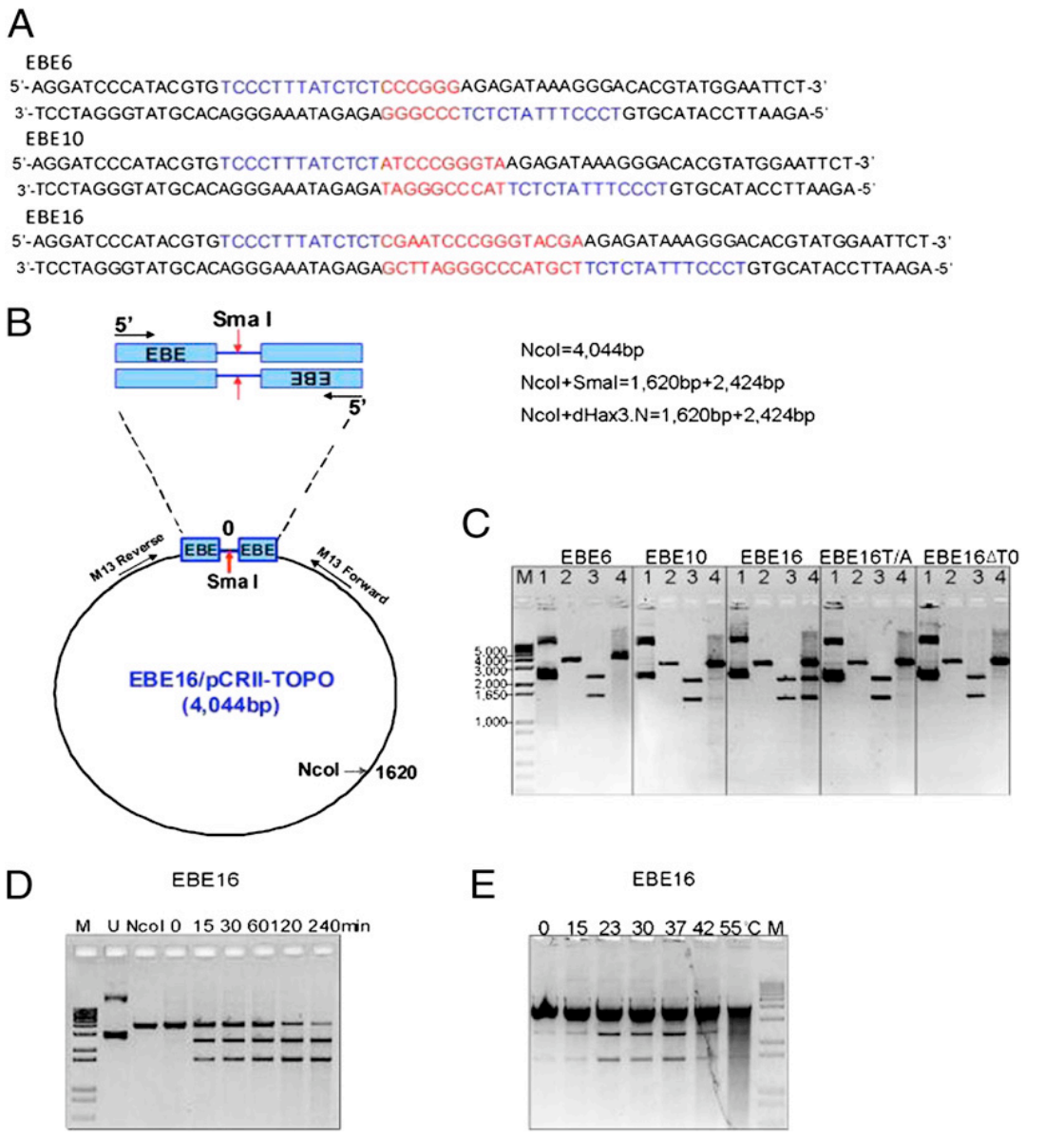


Fig. 5. *In vitro* DNA digestion by dHax3.N Hybrid Nuclease (A) dHax3.N DNA target sequence, EBE sequences in tail-to-tail orientation are shown in blue, and the 6-bp, 10-bp and 16-bp spacer sequences are shown in red. (B) Circular plasmid of pCRII-TOPO containing two EBE target sequence in a tail-to-tail orientation with predicted cleavage sites for SmaI and NcoI enzymes. (C) Activity of dHax3.N Hybrid Nuclease on DNA targets with different spacer lengths. In all panels, M refers to 1-kb Marker, lane 1 is undigested plasmid. Lane 2 is NcoI-linearized plasmid. Lane 3 is plasmid digested with NcoI and SmaI. Lane 4 is NcoI-linearized plasmid digested with dHax3.N for 30 min at 37 °C. Lane 4 in panel EBE6 did not showed the expected cleavage products, indicating that dHax3.N was unable to bind and cut the target sequence with 6-bp spacer

between the two EBEs. Lane 4 in the EBE10 panel showed a little cleavage activity of dHax3.N. Lane 4 of panel EBE16 showed a significant cleavage activity of dHax3.N, where the expected cleavage products can be seen clearly. Lane 4 of panel EBE16 T/A showed very little cleavage activity of dHax3.N, indicating the specificity of dHax3.N. While in the In lane 4 of EBE16ΔT0, where the T nucleotide at the repeat 0 had been removed, there were no cleavage activity of dHax3.N, indicating that the nucleotide T at repeat 0 played an important role in the cleavage activity of dHax3.N. (D) The effect of incubation time on cleavage activity of dHax3.N. (E) The effect of various incubation temperatures on the cleavage activity of dHax3.N (M refers to 1-kb Marker, U refers to unlinearized plasmid, NcoI refers to NcoI-linearized plasmid).

Next we would like to study the binding specificity of dHax3.N in greater detail as well as the effects of mutations on EBE sequence to TALE Nuclease binding and cleavage activity. For that we deleted the nucleotide T at the repeat 0 or replaced T with A in repeat 12 (henceforth these two constructs are called pCRII/EBE16ΔT0 and EBE16T/A, respectively). We observed that nucleotide T deletion in repeat 0 significantly reduced cleavage activity of dHax3.N (Fig. 5C). We also observed that the substitution of T with A at repeat 12 also reduced the cleavage activity of dHax3.N significantly (Fig. 5C). Our results may indicate that the T nucleotide in repeat 0 plays an important role for TALE binding to their target DNA sequence. It also demonstrated that dHax3.N is a very specific endonuclease, in which changes in one nucleotide only in the sequence of the target DNA can inhibit the binding and cleavage activity of dHax3.N significantly.

Furthermore, we also performed cleavage analysis of dHax3.N with varied incubation time and at various temperatures. In general, we observed that the

cleavage products increased with the addition of incubation time (Fig. 5D). For incubation at different temperatures (0° C to 55), our data indicate that the cleavage activity of dHax3.N is most optimal when the digestion reaction was incubated at 23° C and 37° C. Outside this range, the dHax3.N cleavage activity decreased significantly (Fig. 5E). To determine the cleavage sites of dHax3.N, we purified 1.620- and 2.424-bp DNA fragment generated from the cutting of dHax3.N in pCRII/EBE clones for sequencing. The data we obtained confirmed that dHax3.N cut pCRII/EBE clones in the spacer region (Appendix 3).

dHax3.N Hybrid Nuclease Creates DSB *in vivo* in tobacco Transient Assays.

We have demonstrated the ability of dHax3.N to attach and cut DNA in specific target sequences *in vitro*. Further, we would like to study the activity of dHax3.N *in vivo* in plant cells, for that we adapted a modified *A. tumefaciens* mediated tobacco transient assay (Zhang et al., 2010). We constructed two EBE target sequences in a tail-to-tail orientation separated by spacers with the length of 16-bp. A TGA stop codon was inserted in the spacer region and we also inserted an ATG start codon in front of the forward EBE sequence. The ATG.EBE.TGA.spacer.EBE construct was then combined with the GUS-reporter gene (uidA coding sequence) (Fig. 6A). The ATG.EBE.TGA.spacer. EBE::uidA was generated through PCR amplification, and then cloned into pENTR/D vector which then subcloned into pK2GW7 overexpression vector. We hypothesize that

when the construct was introduced and expressed in plant cells, the TGA stop codon would prematurely stop the uidA reporter gene expression.

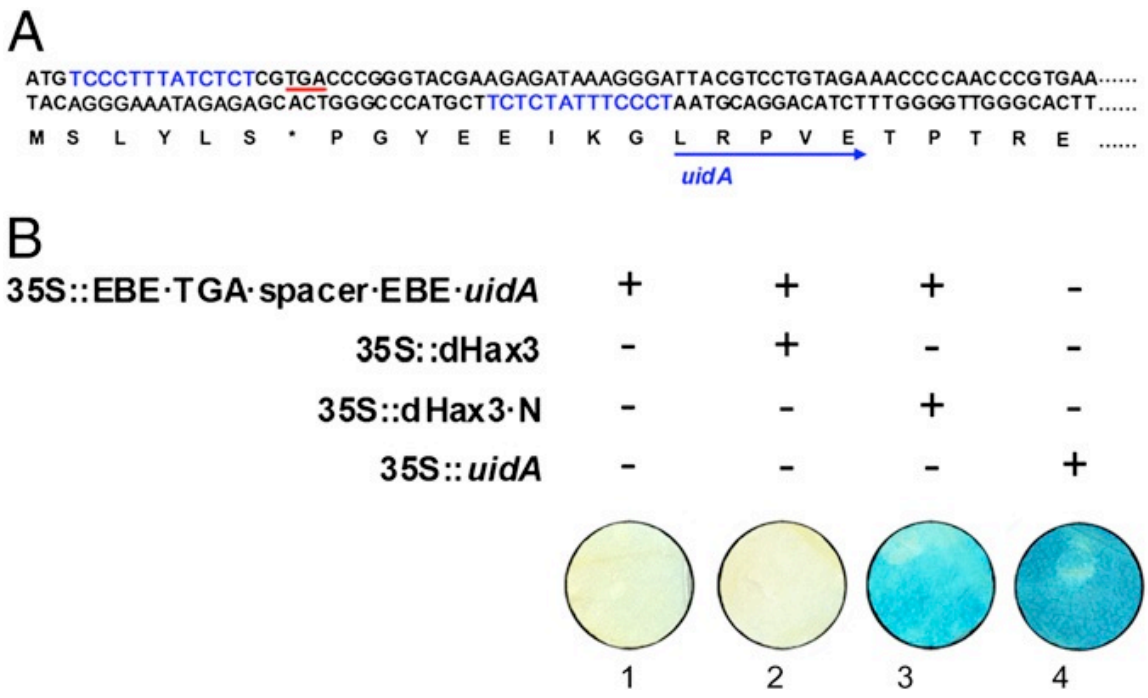


Fig. 6. dHax3.N Hybrid Nuclease induces DSB in vivo. (A) dHax3.N EBE target sequence in a tail-to-tail orientation with a TGA stop codon in the spacer region and a ATG start codon preceding the EBE forward sequence (B) dHax3.N activity in tobacco transient expression assays. Panels 1 and 2 showed no UidA activity in leaves infiltrated with 35S::EBE.TGA.spacer.EBE.uidA or 35S::dHax3, respectively. Panel 3 displayed uidA activity in leaves infiltrated with 35S::EBE.TGA.spacer.EBE.uidA along with 35S::dHax3.N. Panel 4 showed strong uidA activity in leaves infiltrated with 35S::uidA (positive control).

We reasoned that if dHax3.N is expressed and active in plant cells, this protein will cut ATG.EBE.TGA.spacer.EBE::uidA construct in the spacer region and create a DSB. In plant cells, DSBs will be repaired by one of two cellular repair pathways, namely non-homologous end joining (NHEJ) and homologous recombination (HR). NHEJ repair joins the ends of broken strands and results in

small deletions or insertions, which may lead to the elimination of the TGA stop codon and the expression of the uidA reporter. To test this hypothesis, we co-infiltrated or infiltrated separately *A. tumefaciens* carrying 35S::ATG.EBE.TGA.spacer.EBE::uidA, 35S::uidA, 35S::dHax3, and 35S::dHax3.N on *N. benthamiana* leaves.

In accordance with the hypothesis, as shown by GUS staining assay, leaves infiltrated with 35S::ATG.EBE.TGA.spacer.EBE::uidA, 35S::uidA construct showed no UidA reporter activity. UidA reporter activity was also not detected in leaves that are co-infiltrated with 35S::ATG.EBE.TGA.spacer.EBE::uidA and 35S::dHax3 constructs. However, when 35S::ATG.EBE.TGA.spacer.EBE::uidA construct was infiltrated along with 35S::dHax3.N construct, uidA activity is apparent (Fig. 6B). Our results may indicate that the 35S::dHax3.N is expressed and active in plant cells, and able to induce the DSB at the spacer region of 35S::ATG.EBE.TGA.spacer.EBE::uidA construct. Subsequently, the generated DSB was then repaired by cellular repairing machinery which leads to the elimination of TGA stop codon in the spacer region of 35S::ATG.EBE.TGA.spacer.EBE::uidA construct.. As a positive control, 35S::uidA was infiltrated into leaves, where we observed that uidA was clearly expressed (Fig. 4B). Overall, our data demonstrated the TALE-based hybrid nuclease ability to bind to its target and create DSB *in vivo* when transiently expressed in *N. benthamiana* leaves.

IV. DISCUSSION

In this study, we demonstrated the construction of novel TALE hybrid nuclease capable of binding and cutting user pre-defined target sequences, both *in vitro* and *in vivo*. We used Hax3 TALE as the basis for constructing novel hybrid nuclease by combining the DNA binding region of Hax3 with a non-specific cleavage domain from the type II restriction endonuclease *FokI*. We chose Hax3 TALE protein as a template, because this protein has a relatively smaller size than other members of TALEs protein family, but have a long enough central number tandem repeat to recognize and activate specific target DNA sequences (Boch et al., 2009). We showed that our de novo engineered dHax3 and dHax3.N were able to recognize and attach to EBE target sequences *in vitro*, as shown in the results of EMSA. We also showed the ability of our de novo engineered dHax3 and dHax3.N to attach and activate the transcription of the target gene *in vivo*, as shown in the tobacco transient expression assays. It is to be noted that the combination of the *FokI* nuclease domain does not affect the binding specificity of Hax3 TALE as well as the activity of its C-terminal transcription activation domain.

FokI monomers must form a dimer in order to cut the DNA sequences. In ZFNs, two ZFN constructs was designed to be able to identify two target sequences in a tail-to-tail orientation separated by a spacer sequence. When ZFN was attached to the target sequence, the *FokI* can form a dimer and cut the DNA in the spacer

region (Carroll, 2008; Vanamee et al., 2001). Several previous studies have reported that the length of linker between DNA binding modules and nuclease domain, and the spacer length between target sequences significantly influence the cleavage activity of ZFNs. ZFN with a shorter length of linker was found to be able to cut a target with a shorter length of spacer (Bibikova et al., 2001). ZFNs with linker length from 4 to 18-aa showed the highest cleavage activity on the target with the spacer length of 5 to 8-bp (Händel et al., 2008). The length of spacer between the two EBE target sequences necessary for our engineered dHax3.N to show optimal cleavage activity is still not clear.

To test the nuclease activity of dHax3.N and to determine the length of spacer sequence that enables the most efficient dHax3.N cut, we constructed a target site with two EBE target sequences in a tail-to-tail orientation separated by spacers with different lengths. *In vitro* digestion analysis showed that dHax3.N had the highest cleavage activity on the target with 16-bp spacer. The TALE hybrid nuclease that we constructed had a 286-aa long linker between the DNA binding domain and the nuclease domain. The length of the linker in the protein we constructed is likely to be the cause of why our protein had the highest cleavage activity on 16-bp long targets, instead of on targets with 10-bp or 6-bp spacer length. If necessary, we believe it is possible to construct a novel TALE hybrid nuclease capable of recognizing targets with a shorter spacer length by

designing a TALE hybrid nuclease that has a shorter linker between the DNA binding domain and the nuclease domain.

After demonstrating that dHax3.N was able to cut target sequences with the appropriate length of spacer, we further showed the significance of nucleotide T at repeat 0 of the target sequence in the binding and nuclease activity of TALE hybrid nuclease. The removal of nucleotide T at repeat 0 on the target sequence as found in the construct pCRII/EBE16ΔT drastically reduced the cleavage activity of dHax3.N. Additionally, we also showed the specificity of dHax3.N, in which we demonstrated that the replacement of nucleotide T with A at repeat 12 of the target sequence also reduced the cleavage activity of dHax3.N.

Further, we demonstrated the ability of dHax3.N to induce DSB in the target DNA *in vivo* in tobacco transient expression assays. Our data indicate that dHax3.N induced DSB in the spacer region of the target DNA, which was subsequently repaired by the NHEJ repair machinery. DSB repair via the NHEJ pathway caused a small deletion and the elimination of the TGA stop codon at ATG.EBE.TGA.spacer.EBE::uidA constructs, subsequently induced the activation of uidA reporter gene.

The data we obtained showed that TALE hybrid nuclease can be expressed and is active in plant cells; these findings demonstrate the great potential of TALE hybrid nuclease to be used in genome engineering. So far, meganucleases and ZFNs have been widely used to induce site-specific DSBs in the genome of various organisms, but meganucleases and ZFNs technology lack flexibility, adaptability, and reproducibility to be used routinely in genome engineering (Miller et al., 2010; Ramirez et al., 2008). The DNA binding domain of TALEs has clearly predictable sequence specificity, while the specificity of a ZFN designed to recognize a specific target must be identified laboriously from complex expression libraries. Additionally TALEs can be designed to target long DNA target sequences and the modular nature of TALE's tandem-arranged repeats allows rapid construction of novel TALEs with user pre-defined specificity. The data we obtained along with the data published by Li *et al.* (Li et al., 2010; Li et al., 2011), Christian *et al.* (Christian et al., 2010), Cermak *et al.* (Cermak et al., 2011), and Miller *et al.* (Miller et al., 2010) indicate that the DNA binding modules of TALE has the flexibility and adaptability to be used in genome engineering in a simple, reproducible and efficient manner.

In the future, TALE hybrid nuclease can be applied for the purposes of gene disruptions, corrections, and additions. The DNA binding specificity of native or artificial TALE that highly predictable, suggests a variety of applications of this protein. The DNA binding domain of TALEs can be combined with various

functional domains, such as nucleases, methylases, recombinases, or transcription-repressor domains. Several strategies could be applied to mediate TALE hybrid nuclease delivery and expression in plant cells. TALE hybrid nuclease can be transfected into protoplast via chemical transformation or into embryonic culture via gene gun bombardment and chemical transformation. Transformation with *Agrobacterium* or virus-aided gene expression (VAGE) also can be use to introduced TALE hybrid nucleases into the plants cells (Fig. 7).

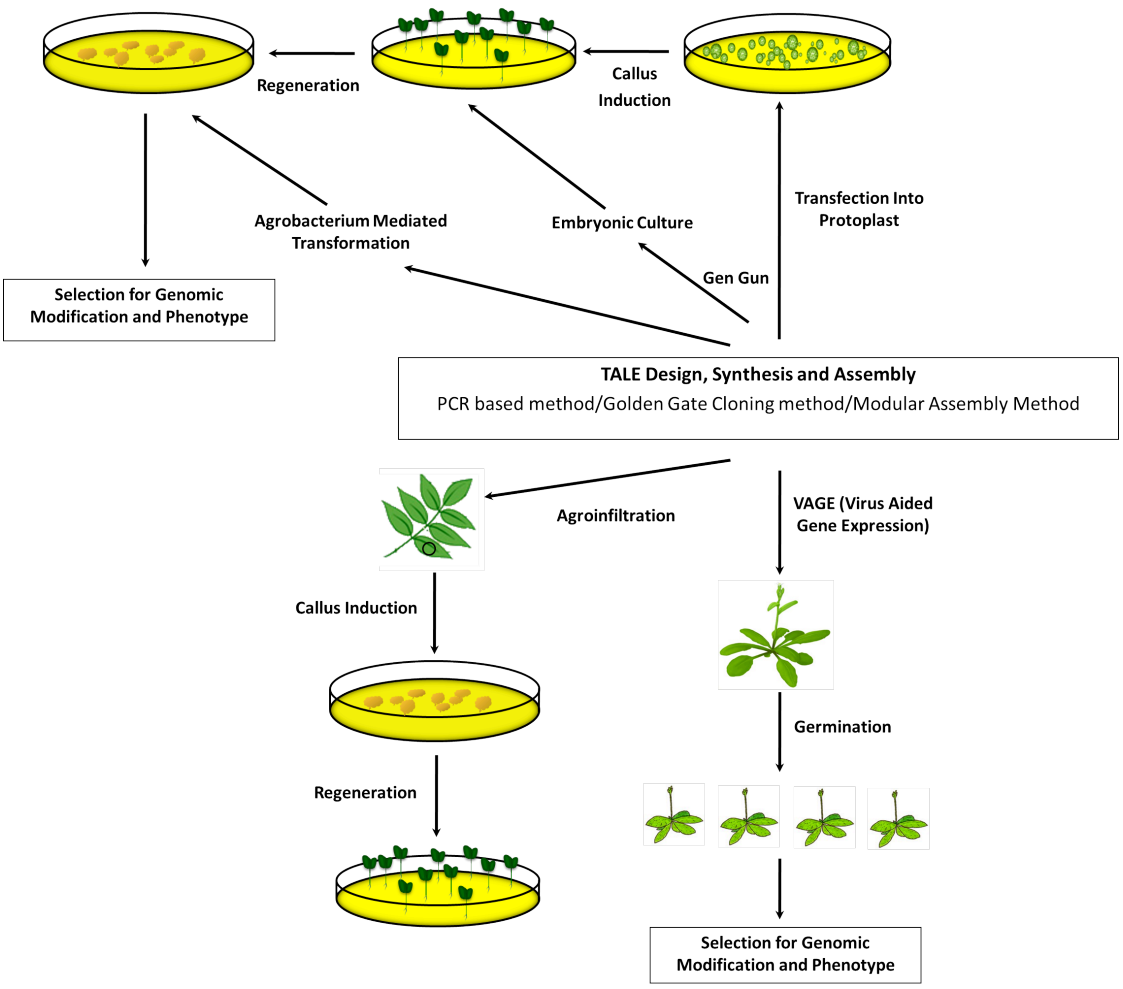


Fig. 7. Strategies for TALE hybrid nucleases delivery and expression in plant cells.

In human cell cultures, Miller *et al.* reported that the efficiency of TALENs in modifying genome target equals that of various ZFNs (Miller et al., 2010). TALEs are still in the initial stage of development and in the future its efficiency can still be increased significantly. Coupled with the simple recognition code of TALE and the modular nature of the TALE DNA recognition code, TALE hold much promise as precise gene editing tools. However, it should be mentioned that there are many unknown details about the nature of binding and the effect of the repeat length and the neighboring repeats. In ZFNs, the interaction between one zinc finger and other zinc fingers as well as the amount of zinc fingers in zinc finger arrays were found to affect the DNA binding activity of zinc finger arrays (Isalan et al., 1997). It should be investigated whether the interaction between the repeat domains in TALE tandem-repeats array and the number of repeats may also affect the DNA binding activity and specificity of TALEs. Additionally, computational studies on positional bias for nucleotides have suggested that there is a strong bias against nucleotides T, A and G at positions 1, 2 and 3 respectively. There is a moderate bias for T in the last half repeat. Consequently there is a corresponding RVD bias: NG and NI are disfavored in repeat 1 and 2 and NN is disfavored in the last half repeat. Several reports have demonstrated a variability of TALEs-mediated transcriptional activation, which might be related to the nature of binding. It should be noted that structural studies on the nature and parameters of TALE-DNA binding is required for effective and efficient design. Finally, in the future it needs to be proven whether the TALE hybrid nuclease can attach to and induce DSB at chromosomal locus. We are currently conducting

experiments to test whether engineered TALE-based hybrid nuclease can recognize and cleave endogenous chromosomal targets. These experiments will be keys to evaluating the application of TALE-based hybrid nuclease for genome engineering in plants and eukaryotes in general. Engineered TALE may significantly advance the genome modification technologies and be a powerful tool for agricultural biotechnology or gene therapy.

BIBLIOGRAPHY

Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G., and Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Molecular and Cellular Biology *21*, 289.

Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. Proceedings of the National Academy of Sciences of the United States of America *95*, 10570.

Boch, J., and Bonas, U. (2010). XanthomonasAvrBs3 Family-Type III Effectors: Discovery and Function. Annual Review of Phytopathology *48*, 419-436.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. Science *326*, 1509-1512.

Bonas, U., Stall, R.E., and Staskawicz, B. (1989). Genetic and structural characterization of the avirulence gene avrBs3 from Xanthomonas campestris pv. vesicatoria. Molecular and General Genetics MGG *218*, 127-136.

Carroll, D. (2008). Progress and prospects: zinc-finger nucleases as gene therapy agents. Gene therapy *15*, 1463-1468.

Cathomen, T., and Joung, J.K. (2008). Zinc-finger nucleases: the next generation emerges. Molecular Therapy *16*, 1200-1207.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Research.

Chiurazzi, M., Ray, A., Viret, J.F., Perera, R., Wang, X.H., Lloyd, A.M., and Signer, E.R. (1996). Enhancement of somatic intrachromosomal homologous recombination in Arabidopsis by the HO endonuclease. The Plant Cell Online *8*, 2057.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). TAL effector nucleases create targeted DNA double-strand breaks. Genetics, 110.120717 v120711.

Collins, C. (2003). Engineering proteins that bind, move, make and break DNA. Current Opinion in Biotechnology *14*, 371-378.

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J.*, et al.* (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nature Biotechnology *26*, 702-708.

Durai, S. (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Research *33*, 5978-5990.

Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M.H., and Chandrasegaran, S. (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Research *33*, 5978.

Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Pâques, F., and Lacroix, E. (2003). A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. Nucleic Acids Research *31*, 2952.

Gepts, P. (2002). A comparison between crop domestication, classical plant breeding, and genetic engineering. Crop Science *42*, 1780.

Gersbach, C.A., Gaj, T., Gordley, R.M., and Barbas, C.F. (2010). Directed evolution of recombinase specificity by split gene reassembly. Nucleic Acids Research *38*, 4198.

Gordley, R.M., Smith, J.D., Gräslund, T., and Barbas Iii, C.F. (2007). Evolution of Programmable Zinc Finger-recombinases with Activity in Human Cells. Journal of Molecular Biology *367*, 802-813.

Gürlebeck, D., Thieme, F., and Bonas, U. (2006). Type III effector proteins from the plant pathogen Xanthomonas and their role in the interaction with the host plant. Journal of Plant Physiology *163*, 233-255.

Händel, E.M., Alwin, S., and Cathomen, T. (2008). Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. Molecular Therapy *17*, 104-111.

Heuer, H., Yin, Y.N., Xue, Q.Y., Smalla, K., and Guo, J.H. (2007). Repeat Domain Diversity of avrBs3-Like Genes in Ralstonia solanacearum Strains and Association with Host Preferences in the Field. Applied and Environmental Microbiology *73*, 4379-4384.

Hopkins, C., White, F., Choi, S., Guo, A., and Leach, J. (1992). Identification of a family of avirulence genes from Xanthomonas oryzae pv. oryzae. Molecular Plant-Microbe Interactions *5*, 451-459.

Isalan, M., Choo, Y., and Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. Proceedings of the National Academy of Sciences of the United States of America *94*, 5617.

Jauhar, P.P. (2001). Genetic engineering and accelerated plant improvement: Opportunities and challenges. Plant Cell *64*, 87-91.

Karimi, M., Depicker, A., and Hilson, P. (2007). Recombinational cloning with plant gateway vectors. Plant physiology *145*, 1144.

Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007). A Bacterial Effector Acts as a Plant Transcription Factor and Induces a Cell Size Regulator. Science *318*, 648-651.

Kay, S., Hahn, S., Marois, E., Wieduwild, R., and Bonas, U. (2009). Detailed analysis of the DNA recognition motifs of theXanthomonastype III effectors AvrBs3 and AvrBs3Δrep16. The Plant Journal *59*, 859-871.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences *93*, 1156.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (2010). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Research *39*, 359-372.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P., and Yang, B. (2011). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Research.

Lloyd, A. (2005). Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proceedings of the National Academy of Sciences *102*, 2232-2237.

Mani, M., Kandavelou, K., Dy, F.J., Durai, S., and Chandrasegaran, S. (2005). Design, engineering, and characterization of zinc finger nucleases. Biochemical and Biophysical Research Communications *335*, 447-457.

Marois, E., Van den Ackerveken, G., and Bonas, U. (2002). The Xanthomonas type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host. Molecular Plant-Microbe Interactions *15*, 637-646.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J.*, et al.* (2010). A TALE nuclease architecture for efficient genome editing. Nature Biotechnology *29*, 143-148.

Minsavage, G., Dahlbeck, D., Whalen, M., Kearney, B., Bonas, U., Staskawicz, B., and Stall, R. (1990). Gene-for-gene relationships specifying disease resistance in Xanthomonas campestris pv. vesicatoria--pepper interactions. Molecular Plant-Microbe Interactions *3*, 41-47.

Morbitzer, R., Elsaesser, J., Hausner, J., and Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Research.

Morbitzer, R., Romer, P., Boch, J., and Lahaye, T. (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proceedings of the National Academy of Sciences *107*, 21617-21622.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Pabo, C.O., Peisach, E., and Grant, R.A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. Annual Review of Biochemistry *70*, 313-340.

Ramalingam, S., Kandavelou, K., Rajenderan, R., and Chandrasegaran, S. (2010). Creating Designed Zinc Finger Nucleases with Minimal Cytotoxicity. Journal of Molecular Biology.

Ramirez, C.L., Foley, J.E., Wright, D.A., Müller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., and Townsend, J.A. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. Nature methods *5*, 374-375.

Thermes, V., Grabher, C., Ristoratore, F., Bourrat, F., Choulika, A., Wittbrodt, J., and Joly, J.S. (2002). I-SceI meganuclease mediates highly efficient transgenesis in fish. Mechanisms of Development *118*, 91-98.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nature Reviews Genetics *11*, 636-646.

Vanamee, É.S., Berriman, J., and Aggarwal, A.K. (2007). An EM view of the FokI synaptic complex by single particle analysis. Journal of Molecular Biology *370*, 207-212.

Vanamee, É.S., Santagata, S., and Aggarwal, A.K. (2001). FokI requires two specific DNA sites for cleavage1. Journal of Molecular Biology *309*, 69-78.

Wah, D.A., Bitinaite, J., Schildkraut, I., and Aggarwal, A.K. (1998). Structure of FokI has implications for DNA cleavage. Proceedings of the National Academy of Sciences of the United States of America *95*, 10564.

Wang, H., Perrault, A.R., Takeda, Y., Qin, W., and Iliakis, G. (2003). Biochemical evidence for Ku independent backup pathways of NHEJ. Nucleic Acids Research *31*, 5377.

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature *460*, 894-898.

Weber, E., Gruetzner, R., Werner, S., Engler, C., and Marillonnet, S. (2011). Assembly of Designer TAL Effectors by Golden Gate Cloning. PloS One *6*, e19722.

Wirth, D., Gamanorton, L., Riemer, P., Sandhu, U., Schucht, R., and Hauser, H. (2007). Road to precision: recombinase-based targeting technologies for genome engineering. Current Opinion in Biotechnology *18*, 411-419.

Yamaguchi-Shinozaki, K., and Shinozaki, K. (1993). Characterization of the expression of a desiccation-responsive rd29 gene of Arabidopsis thaliana and analysis of its promoter in transgenic plants. Molecular and General Genetics MGG *236*, 331-340.

Yang, B., and White, F.F. (2004). Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. Molecular Plant-Microbe Interactions *17*, 1192-1200.

Zhang, F. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nature Biotechnology.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X., Pierick, C.J., Dobbs, D., and Peterson, T. (2010). High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proceedings of the National Academy of Sciences *107*, 12028.

Zhong, G.Y. (2001). Genetic issues and pitfalls in transgenic plant breeding. Euphytica *118*, 137-144.

APPENDICES

Appendix 1. dHax3 and dHax3.N Sequence

dHax3 Sequence

ATGGACCCAATACGAAGCAGAACGCCATCACCAGCTAGGGAACTTCTCTCTGGACCACA
GCCTGATGGAGTTCAGCCAACTGCAGATCGAGGTGTTTCTCCGCCAGCCGGTGGCCCTT
TAGATGGTCTCCCAGCAAGAAGAACAATGTCCCGTACCAGACTCCCAAGTCCCCCTGCC
CCGTCGCCAGCCTTTTCAGCTGACTCCTTCTCTGATCTTCTTAGGCAATTTGACCCTTC
TCTTTTCAATACATCCCTTTTCGATTCACTTCCTCCTTTCGGCGCACATCATACTGAGG
CAGCCACCGGCGAATGGGACGAAGTCCAAAGTGGTTTAAGGGCAGCTGATGCTCCACCA
CCGACGATGAGAGTCGCTGTTACCGCCGCACGTCCTCCTAGAGCCAAGCCAGCCCCTAG
AAGACGAGCTGCGCAACCCTCCGATGCAAGCCCTGCAGCTCAAGTAGACCTTCGAACAC
TAGGTTACTCCCAGCAACAACAAGAAAAAATAAAGCCAAAGGTTAGATCTACAGTTGCA
CAACATCACGAAGCCCTAGTCGGACACGGATTTACACATGCTCATATCGTGGCTCTTTC
ACAACATCCTGCAGCTCTTGGAACAGTCGCTGTCAAATATCAGGATATGATTGCTGCAT
TGCCAGAAGCTACTCACGAAGCTATCGTCGGAGTTGGGAAACAATGGTCAGGCGCAAGA
GCATTAGAGGCGCTTCTCACCGTAGCTGGTGAATTACGAGGTCCTCCACTCCAATTGGA
TACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCGAAGCCGTTC
ATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTTACTCCTGAACAGGTTGTC
GCAATAGCTTCACATGATGGCGGAAAACAAGCTCTTGAAACAGTGCAACGTCTCCTTCC
CGTCCTCTGTCAGGCTCACGGATTGACTCCTCAGCAGGTCGTCGCAATTGCATCACATG
ATGGAGGCAAACAAGCTTTAGAAACAGTACAAAGACTATTGCCCGTTCTTTGCCAAGCG
CATGGGTTAACTCCCGAACAAGTCGTTGCCATTGCAAGTCACGACGGAGGTAAACAAGC
TCTCGAAACGGTTCAAGCACTTTTACCCGTTCTCTGTCAAGCACATGGACTCACACCTG
AACAAGTAGTTGCTATCGCATCGAATGGAGGTGGAAAACAAGCACTGGAAACTGTACAA
AGACTTTTGCCAGTTTTATGTCAAGCGCACGGTCTTACTCCTCAACAAGTTGTCGCCAT
TGCCTCTAACGGTGGTGGAAAACAAGCTCTTGAAACTGTCCAGAGACTTCTGCCCGTTC
TATGTCAGGCTCATGGGCTAACCCCTCAACAGGTTGTTGCAATCGCATCTAATGGAGGA
GGAAAACAAGCTTTAGAAACTGTCCAACGACTACTGCCCGTTCTCTGCCAAGCACACGG
ACTTACCCCACAACAAGTTGTGGCAATAGCTTCTAATTCTGGTGGTAAACAAGCCCTTG
AGACGGTTCAAAGACTTCTACCAGTTCTTTGTCAGGCACATGGATTGACCCCACAACAG
GTCGTAGCAATCGCATCTAATGGAGGTGGTAAGCAAGCTCTAGAAACGGTACAAAGATT
ACTTCCCGTGCTTTGTCAAGCTCATGGACTCACTCCTCAACAAGTGGTCGCTATTGCAA
GTCATGATGGTGGAAAGCAAGCACTAGAAACCGTCCAACGACTCCTTCCTGTTCTCTGT
CAAGCACATGGTCTTACGCCCGAACAAGTTGTTGCTATAGCTTCGAACGGAGGTGGAAA
ACAAGCTCTCGAAACCGTCCAAAGGCTCCTCCCAGTACTTTGCCAAGCACATGGATTAA
CCCCTGAGCAAGTAGTTGCAATTGCCTCGCACGACGGAGGAAAGCAAGCATTAGAAACT
GTTCAGAGACTTTTGCCTGTCCTGTGTCAAGCCCACGGTCTAACACCACAACAAGTCGT
CGCAATCGCTAGTAATGGAGGAGGTAGACCTGCATTGGAGTCGATAGTCGCACAACTAT
CACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAGTCGCACTTGCATGT

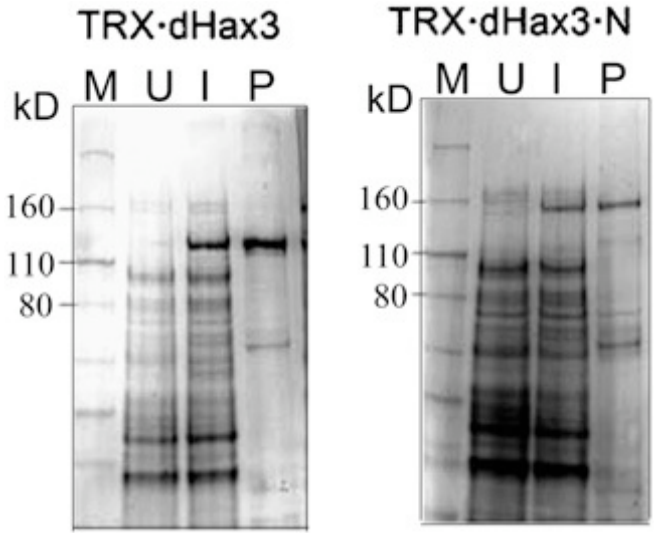
TTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCATT
GATTAAAAGAACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATG
CTCAAGTCGTAAGAGTTTTGGGTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTT
GACGATGCAATGACTCAATTTGGAATGAGTAGACATGGACTCCTGCAATTATTTCGAAG
GGTCGGAGTTACAGAGCTCGAAGCCAGGTCAGGAACGCTGCCCCCCGCATCTCAACGAT
GGGATAGAATTCTCCAAGCCTCTGGAATGAAAAGAGCTAAACCTTCACCAACGTCCACA
CAAACACCAGACCAAGCTTCTCTCCACGCTTTTGCCGACTCACTAGAGAGAGATCTAGA
TGCACCGTCACCTATGCATGAAGGAGACCAAACAAGAGCCTCTTCAAGAAAACGTTCTC
GTTCTGATAGAGCTGTCACTGGACCTTCCGCCCAACAATCTTTCGAAGTCCGAGTTCCT
GAGCAACGAGATGCCCTACACCTGCCTTTGCTTTCTTGGGGAGTTAAGCGACCACGTAC
TAGAATTGGTGGACTACTCGATCCAGGTACACCAATGGATGCTGATCTCGTTGCTTCCT
CTACCGTAGTATGGGAGCAAGACGCAGACCCCTTCGCTGGAACTGCTGACGATTTCCCA
GCCTTTAACGAGGAAGAATTGGCTTGGTTAATGGAACTTCTACCGCAATGA

dHax3.N sequence

ATGGACCCAATACGAAGCAGAACGCCATCACCAGCTAGGGAACTTCTCTCTGGACCACA
GCCTGATGGAGTTCAGCCAACTGCAGATCGAGGTGTTTCTCCGCCAGCCGGTGGCCCTT
TAGATGGTCTCCCAGCAAGAAGAACAATGTCCCGTACCAGACTCCCAAGTCCCCCTGCC
CCGTCGCCAGCCTTTTCAGCTGACTCCTTCTCTGATCTTCTTAGGCAATTTGACCCTTC
TCTTTTCAATACATCCCTTTTCGATTCACTTCCTCCTTTCGGCGCACATCATACTGAGG
CAGCCACCGGCGAATGGGACGAAGTCCAAAGTGGTTTAAGGGCAGCTGATGCTCCACCA
CCGACGATGAGAGTCGCTGTTACCGCCGCACGTCCTCCTAGAGCCAAGCCAGCCCCTAG
AAGACGAGCTGCGCAACCCTCCGATGCAAGCCCTGCAGCTCAAGTAGACCTTCGAACAC
TAGGTTACTCCCAGCAACAACAAGAAAAAATAAAGCCAAAGGTTAGATCTACAGTTGCA
CAACATCACGAAGCCCTAGTCGGACACGGATTTACACATGCTCATATCGTGGCTCTTTC
ACAACATCCTGCAGCTCTTGGAACAGTCGCTGTCAAATATCAGGATATGATTGCTGCAT
TGCCAGAAGCTACTCACGAAGCTATCGTCGGAGTTGGGAAACAATGGTCAGGCGCAAGA
GCATTAGAGGCGCTTCTCACCGTAGCTGGTGAATTACGAGGTCCTCCACTCCAATTGGA
TACTGGGCAATTATTAAAAATCGCTAAACGAGGTGGAGTCACTGCTGTCGAAGCCGTTC
ATGCATGGCGTAACGCTCTCACGGGCGCACCACTAAACCTTACTCCTGAACAGGTTGTC
GCAATAGCTTCACATGATGGCGGAAAACAAGCTCTTGAAACAGTGCAACGTCTCCTTCC
CGTCCTCTGTCAGGCTCACGGATTGACTCCTCAGCAGGTCGTCGCAATTGCATCACATG
ATGGAGGCAAACAAGCTTTAGAAACAGTACAAAGACTATTGCCCGTTCTTTGCCAAGCG
CATGGGTTAACTCCCGAACAAGTCGTTGCCATTGCAAGTCACGACGGAGGTAAACAAGC
TCTCGAAACGGTTCAAGCACTTTTACCCGTTCTCTGTCAAGCACATGGACTCACACCTG
AACAAGTAGTTGCTATCGCATCGAATGGAGGTGGAAAACAAGCACTGGAAACTGTACAA
AGACTTTTGCCAGTTTTATGTCAAGCGCACGGTCTTACTCCTCAACAAGTTGTCGCCAT
TGCCTCTAACGGTGGTGGAAAACAAGCTCTTGAAACTGTCCAGAGACTTCTGCCCGTTC
TATGTCAGGCTCATGGGCTAACCCCTCAACAGGTTGTTGCAATCGCATCTAATGGAGGA
GGAAAACAAGCTTTAGAAACTGTCCAACGACTACTGCCCGTTCTCTGCCAAGCACACGG
ACTTACCCCACAACAAGTTGTGGCAATAGCTTCTAATTCTGGTGGTAAACAAGCCCTTG
AGACGGTTCAAAGACTTCTACCAGTTCTTTGTCAGGCACATGGATTGACCCCACAACAG
GTCGTAGCAATCGCATCTAATGGAGGTGGTAAGCAAGCTCTAGAAACGGTACAAAGATT
ACTTCCCGTGCTTTGTCAAGCTCATGGACTCACTCCTCAACAAGTGGTCGCTATTGCAA

GTCATGATGGTGGAAAGCAAGCACTAGAAACCGTCCAACGACTCCTTCCTGTTCTCTGT
CAAGCACATGGTCTTACGCCCGAACAAGTTGTTGCTATAGCTTCGAACGGAGGTGGAAA
ACAAGCTCTCGAAACCGTCCAAAGGCTCCTCCCAGTACTTTGCCAAGCACATGGATTAA
CCCCTGAGCAAGTAGTTGCAATTGCCTCGCACGACGGAGGAAAGCAAGCATTAGAAACT
GTTCAGAGACTTTTGCCTGTCCTGTGTCAAGCCCACGGTCTAACACCACAACAAGTCGT
CGCAATCGCTAGTAATGGAGGAGGTAGACCTGCATTGGAGTCGATAGTCGCACAACTAT
CACGACCTGATCCCGCTCTTGCAGCATTGACAAACGATCATTTAGTCGCACTTGCATGT
TTAGGAGGACGACCAGCACTTGATGCCGTTAAGAAAGGACTACCGCACGCCCCTGCATT
GATTAAAAGAACAAACAGACGAATCCCGGAGAGAACTTCACATCGTGTAGCCGATCATG
CTCAAGTCGTAAGAGTTTTGGGTTTCTTCCAATGTCATTCCCACCCAGCTCAAGCTTTT
GACGATGCAATGACTCAATTTGGAATGAGTAGACATGGACTCCTGCAATTATTTCGAAG
GGTCGGAGTTACAGAGCTCGAAGCCAGGTCAGGAACGCTGCCCCCCGCATCTCAACGAT
GGGATAGAATTCTCCAAGCCTCTGGAATGAAAAGAGCTAAACCTTCACCAACGTCCACA
CAAACACCAGACCAAGCTTCTCTCCACGCTTTTGCCGACTCACTAGAGAGAGATCTAGA
TGCACCGTCACCTATGCATGAAGGAGACCAAACAAGAGCCTCTTCAAGAAAACGTTCTC
GTTCTGATAGAGCTGTCACTGGACCTTCCGCCCAACAATCTTTCGAAGTCCGAGTTCCT
GAGCAACGAGATGCCCTACACCTGCCTTTGCTTTCTTGGGGAGTTAAGCGACCACGTAC
TAGAATTGGTGGACTACTCGATCCAGGTACACCAATGGATGCTGATCTCGTTGCTTCCT
CTACCGTAGTATGGGAGCAAGACGCAGACCCCTTCGCTGGAACTGCTGACGATTTCCCA
GCCTTTAACGAGGAAGAATTGGCTTGGTTAATGGAACTTCTACCGCAAAAGCAACTAGT
CAAAAGTGAACTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTC
ATGAATATATTGAATTAATTGAAATTGCCAGAAATTCCACTCAGGATAGAATTCTTGAA
ATGAAGGTAATGGAATTTTTTATGAAAGTTTATGGATATAGAGGTAAACATTTGGGTGG
ATCAAGGAAACCGGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTGA
TCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAA
ATGCAACGATATGTCGAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATG
GTGGAAAGTCTATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACT
TTAAAGGAAACTACAAAGCTCAGCTTACACGATTAAATCATATCACTAATTGTAATGGA
GCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATT
AACCTTAGAGGAAGTGAGACGGAAATTTAATAACGGCGAGATAAACTTTTGA

Appendix 2. TRX.dHax3 and TRX.dHax3.N proteins migrate to their expected sizes on SDS gel electrophoresis. M, protein marker; U, uninduced fraction; I, induced fraction (IPTG was used as the inducer); P, purified protein.



Appendix 3. DNA sequencing of cleavage products confirm cleavage sites within the spacer regio